Experimental Procedures

p66 subunit of HIV-1 reverse transcriptase. Expression, purification, deuteration and nitroxide spin-labeling of the p66 subunit of HIV-1 reverse transcriptase was carried out as described previously.[1] In this construct all natural cysteines were mutated to Ala, and nitroxide spin-labels were attached to either G504C or Q547C located within the RNase H domain.

HIV-1 protease group O (PR). Expression, purification, deuteration and nitroxide spin-labeling of PR was as described previously.[2] DEER measurements were carried out on a double mutant bearing the active site D25N mutation to inactivate the protease, and the V82C mutation for nitroxide spin-labeling. PR from group O HIV-1 was chosen over that from group M as group O PR is soluble at pH 7-8, the pH solubility range for RT p66/p66', while group M PR becomes insoluble above pH 6.[2]

DEER sample preparation. Nitroxide (R1) spin-labeling was carried out with S-(1-oxyl-2,2,5,5-tetramethyl-2,5-dihydro-1H-pyrrol-3-yl)methyl methanesulfonothioate (MTSL; Toronto Research Chemicals) as described previously.[1, 3] Nitroxide-labeling was verified by electrospray ionization mass spectrometry. (Note that even if nitroxide labeling is incomplete, this will only impact the modulation depth, and other than issues associated with a concomitant reduction in signal-to-noise, will have no effect on the resulting $P(r)$ distance distributions or the accuracy of the mean distances.) The buffer for all DEER samples (referred to hereafter as DEER buffer) comprised 25 mM Tris-HCl, pH 7.6, 300 mM NaCl, 20 mM $MgCl_2$, 30% (w/v) deuterated (1,1,2,3,3-d5, 99 atom % D) glycerol, and 70% (v/v) D_2O (99.9 atom % D). Samples (12 µl) for DEER were pipetted into 1 mm inner diameter/ 1.6 mm outer diameter quartz EPR tubes (VitroCom Inc.) and flash frozen in liquid nitrogen.

HIV-1 protease inhibitors. Stock solutions (100 mM in DMSO) of the active site inhibitors darunavir (DRV, NIH AIDS Research and Reference Reagent Program), DMP-323,[4] and the non-hydrolyzable substrate analogue inhibitor RPB,[5] Ac-Thr-Ile-Nle-r-Nle-Gln-Arg-NH_2, where r (reduced peptide bond) is a methylene group that replaces the peptide carbonyl (Bachem Bioscience Inc.) and Nle is norleucine, were diluted to 0.3 mM with 5 mM sodium acetate, pH 6, and then to the appropriate concentration with DEER buffer. The S4 peptide (Lys-Ala-Arg-Val-Nle-[4-nitrophenylalanine]-Glu-Ala-Nle-NH2,[6-7] (California Peptide Research, Napa, CA) was initially dissolved in H_2O to make a 4.3 mM stock and then further diluted with DEER buffer to make a 0.3 mM sub-stock.

Wild type protease O activity. The activity of wild type mature protease O (PR) was determined in 50 mM sodium acetate buffer, pH 5, and 250 mM NaCl at a final concentration of 0.41 µM PR and 380 µM of the chromogenic substrate S4 in a total volume of 120 µl at 28 ºC.[7] Buffers for assays at pH 6.5 and 7.5 contained 25 mM sodium phosphate and 300 mM NaCl. At pH 5, the values of k_{cat} and K_m for the chromogenic substrate are 240 ± 10 min^{-1} and 71 ± 12 µM, respectively. The activity of PR is minimally affected by the addition of 30% (v/v) glycerol in the assay. PR-O activity is reduced as the pH is increased, becoming negligible at pH 7.5, consistent with the pH-rate profile.[8] The affinity of wild type PR-O for darunivir is 11±8 nM, carried out in 50 mM Tris-HCl, pH 7.5 and 50 mM NaCl at 28°C, as determined by isothermal titration calorimetry using an iTC200 microcalorimeter (Malvern Instruments Inc., Westborough, MA). Wild type PR was used as a template to generate the mutants, D25N, and D25N/V82C. These mutants, also used in NMR studies described previously,[2] display excellent ^{1}H-^{15}N correlation spectra characteristic of the stable PR dimeric fold: the ^{1}H/^{15}N cross-peaks for PR (D25N/V81C) are only perturbed in the immediate vicinity of the V81C mutation; and paramagnetic relaxation enhancement data on PR(V82C-R1) are fully consistent with the known structure of PR, with backbone amide protons in close proximity to V81C-R1 broadened as a consequence of dipolar interactions with the unpaired electron on the nitroxide R1 spin label (L. Deshmukh, J.M.L. and G.M.C., unpublished data).[2] It should be noted that both PR from HIV-1 groups O and M, as well as retroviral proteases in general, have a broad range of specificities. It is therefore difficult to assess the exact determinants of PR substrate specificity or to derive a generalized consensus sequence(s) recognized by PR; however, fit complementarity and multiple intermolecular contacts between substrate and the catalytic cleft and flaps of PR are important in increasing the lifetime of the complex and therefore the likelihood of proteolysis.[9]

Pulsed Q-band EPR spectroscopy. Pulsed EPR data were collected at Q-band (33.8 GHz) and 50 K on a Bruker E-580 spectrometer equipped with a 150 W traveling-wave tube (TWT) amplifier, a model ER5107D2 resonator, and a cryofree cooling unit, as described previously.[1] DEER experiments were acquired using a conventional four-pulse sequence.[10] The observer and ELDOR pump pulses were separated by ca. 90 MHz with the observe $\pi/2$ and π pulses set to 12 and 24 ns, respectively, and the ELDOR π pulse to 10 ns. The π observer pulse of 24 ns may suppress to some extent shorter distances less than 15 Å due to insufficient excitation bandwidth. The pump frequency was centered at the Q-band nitroxide spectrum located at +40 MHz from the center of the resonator frequency. The τ_1 value of 400 ns for the first echo-period time was incremented eight times in 16 ns steps to average ^{2}H modulation. To optimize data point sampling, the position of the ELDOR pump pulse was incremented in steps of $\Delta t = 4$ ns for short distances (Fig. 2) and 12-24 ns for long distances (Fig. 3). Data acquisition for each DEER echo curve was carried out for 12-24 h with between 13-32 scans using a shot repetition time of 2.5 μs and a shot number of 150 scans per point. The bandwidth of the overcoupled resonator was 120 MHz. The second echo period time τ_2 was set to t_{max}+ 700 ns, where t_{max} is the maximum dipolar acquisition time; data collection was not extended to the full τ_2 range because of a persistent "2+1" echo perturbation of the DEER echo curves at a time of about τ_1 from the final observe π pulse.

***P(r)* distributions from DEER data.** DEER echo curves were fitted using three approaches to generate *P*(*r*) distributions: (a) the program DD[11] which uses a sum of Gaussians to directly fit the experimental DEER data (including automated background correction with a best-fit exponential decay); (b) validated Tikhonov regularization using DeerAnalysis 2016;[12] and (c) the program WavPDS which uses a wavelet approach to filter out noise and obtains P(*r*) distributions by singular value decomposition (SVD).[13-14] The optimal number of Gaussians used by the program DD to represent the DEER data was assessed using the Akaike information criterion corrected for finite sample size (AICc) and the Bayesian information criterion (BIC).[11] Validated Tikhonov regularization was carried out varying the modulation depth (11 steps), background density (11 steps), background start (11 steps) and noise sensitivity (5 steps) for a total of 6655 permutations. The Tikhonov regularization parameter α was automatically determined by DeerAnalysis for each iteration. The WavPDS Method uses the Daubechis 6 wavelet family to filter out noise; post 'denoising', the DEER echo curves were analysed by SVD reconstruction. The SVD cutoff was assessed from a Picard plot.[14] To derive mean distances and populations, the *P*(*r*) distributions obtained from DA and SVD were further analyzed by fitting them to a sum of Gaussians using the same number used for DD.

Modeling of *P*(r) distributions based on atomic coordinates. *P*(*r*) distributions from atomic coordinates were calculated using two alternate approaches: the rotamer library-based program MMM[15-16] and RosettaEPR.[17]

In the case of MMM,[15-16] two sets of *P*(*r*) distributions were calculated: the raw distribution making use of all generated rotamer conformers and the conformer population optimized distribution using the feature ANY_ROTAMER. In the latter, the individual conformers within the raw ensemble are weighted to best reproduce the experimental *P*(*r*) distribution.

RosettaEPR[17] modeling was carried out as follows. A 100 conformers of MTSL, attached to V82C of both subunits, were generated using the fixed backbone application of Rosetta[18-19], based on a MTSL rotamer library of 54 conformations derived from high resolution crystal structures. These calculations were carried out for free PR in the closed (1F7A with S4 substrate removed) and open (1TW7) flap configurations and for PR complexed to the S4 peptide (1F7A). The latter was used for all drug and peptide complexes as the version of Rosetta we employed is not parameterized for the various drug molecules or for a peptide such as RPB which has a reduced peptide bond. The rotamer ensemble was then screened against the experimental *P*(*r*) distribution using a genetic algorithm to identify MTSL rotamers that reproduce the experimental distances. An initial ensemble of 5 models was randomly generated from the pool of 100 models. In each generation of the genetic algorithm, a model can be added, removed or swapped for another model. Each new ensemble (ranging from 5 up to 11 members) was evaluated based on the sum of the cumulative Euclidian distance between the ensemble back-calculated *P*(*r*) distribution and experimental one. If the cumulative Euclidean distance is smaller than that of the previous generation ensemble, the new ensemble is selected. A total of 4000 generations was performed.

Author Contributions

T.S., and GMC designed research; T.S., J.M.L. and G.M.C. performed research; T.S. and G.M.C. analyzed data; T.S. and G.M.C. wrote paper.

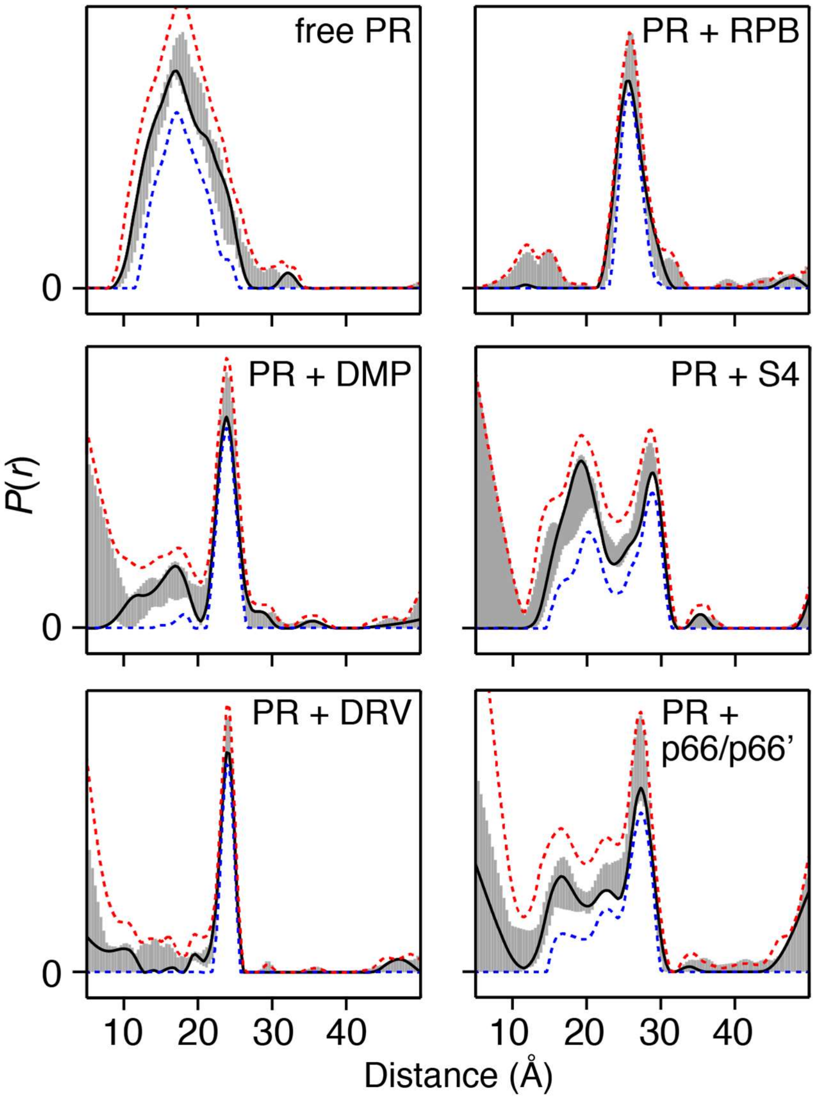

Figure S1. Results of DeerAnalysis[12] validation of *P*(*r*) distributions obtained using Tikhonov regularization for the PR (V82C-R1) dimer free and in the presence of drugs and substrates. The *P*(*r*) distribution with the best r.m.s.d. is displayed in black; the grey regions indicate the full *P*(*r*) variation over all trials; and the red and blue dotted lines are the upper (mean value probability plus two times its S.D.) and lower (mean value minus two times its S.D.) error estimates, respectively.

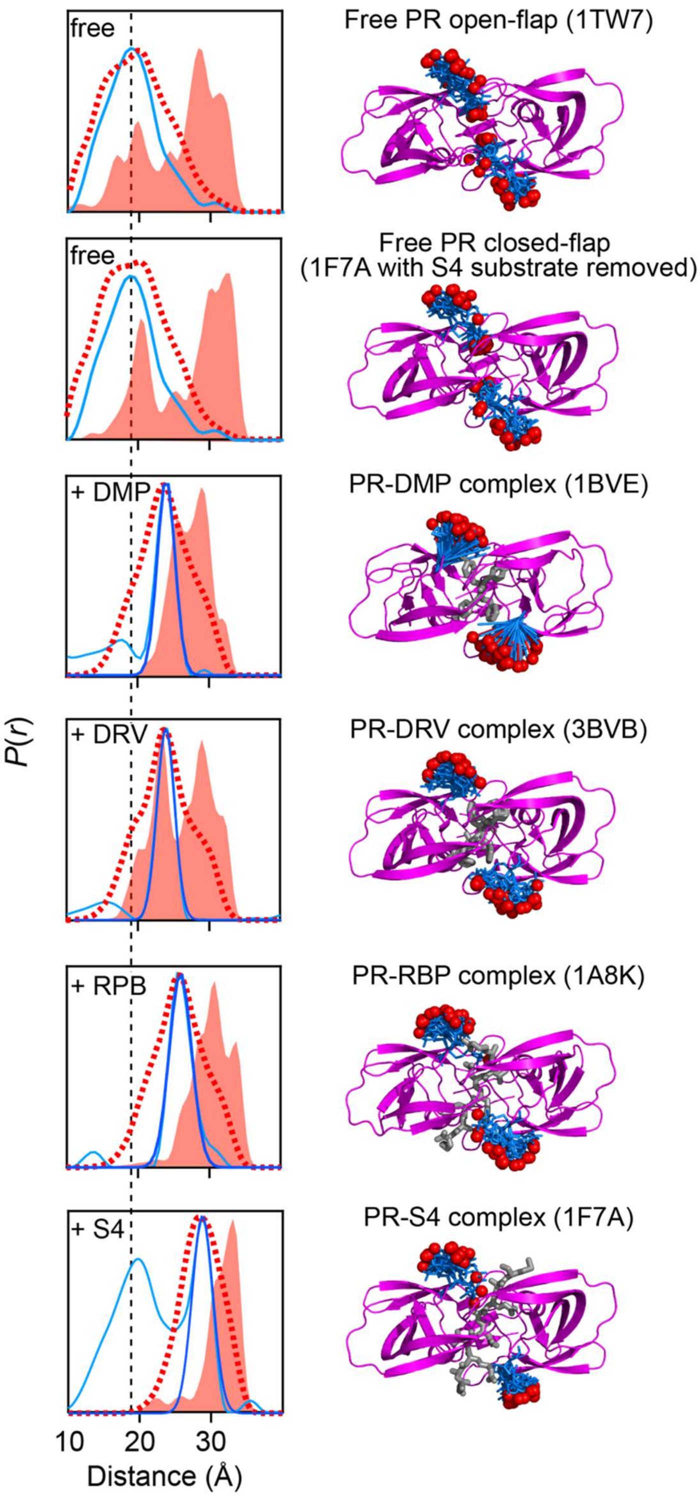

Figure S2. Comparison of experimental DEER-derived *P*(*r*) distributions for nitroxide-labeled HIV-1 protease (V82C-R1) free and in the presence of drug inhibitors and peptides with the raw and rotamer population optimized *P*(*r*) distributions calculated from the atomic coordinates using the program MMM.[15] The experimental DEER-derived *P*(*r*) distributions obtained by Tikhonov regularization using DeerAnalysis[12] are shown in light blue; for the complexes, a Gaussian shown in dark blue was fitted to the peak corresponding to the bound state; the shaded red areas and dashed red lines represent the raw and conformer population optimized *P*(*r*) distributions, respectively, calculated from the coordinates using MMM.[16] The structures of PR (purple ribbons) showing the raw MMM ensemble of nitroxide MTSL label conformations (blue sticks with the oxygen bearing the unpaired electron shown as red balls) are shown on the right; the drugs and peptide substrates are shown as grey sticks. For free PR, the MMM calculations were performed using the coordinates of the open flap conformation (PDB code 1TW7) as well as the closed flap conformation with the S4 peptide substrate removed (PDB code 1F7A).[20] The predominant flap conformation of PR (both groups O and M) in free solution is in the closed state as deduced from residual dipolar coupling (RDC) measurements[9, 21] Also note that aside from the flap region, there is minimal variation in backbone coordinates both in solution and in crystal structures, and the structure of the PR backbone in solution is fully consistent with that of high resolution crystal structures as judged by RDCs.[9, 21] Thus, it is perfectly reasonable to assume that the position of the backbone around the V82C-R1 nitroxide spin label in solution, both free and in the various complexes, is faithfully represented by the crystal structures.

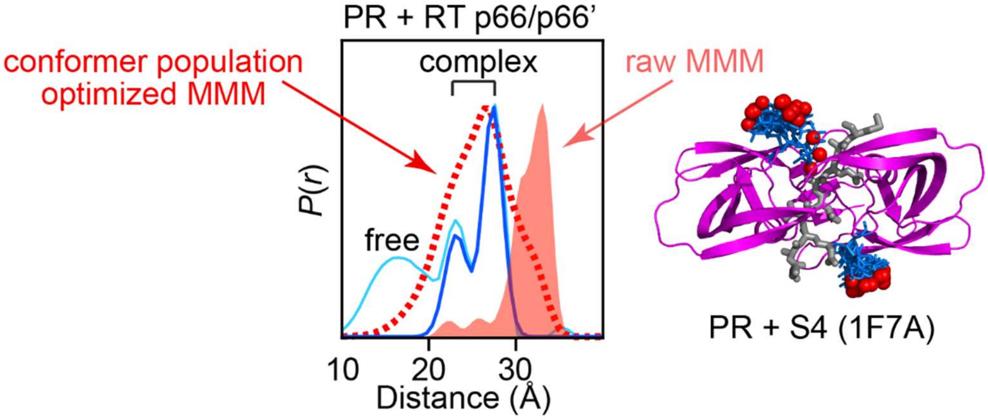

Figure S3. Comparison of the experimental DEER-derived *P*(*r*) distribution for nitroxide-labeled HIV-1 protease (V82C-R1) in the presence of the RT p66/p66' precursor with the raw and rotamer optimized *P*(*r*) distributions calculated from the atomic coordinates using the program MMM.[15] The experimental DEER-derived *P*(*r*) distribution was obtained by Tikhonov regularization using DeerAnalysis[12] and is shown in light blue; two Gaussians shown in dark blue were fitted to the two peaks corresponding to the bound state; the shaded red areas and dashed red lines represent the raw and conformer population optimized *P*(*r*) distributions, respectively, calculated from the coordinates using MMM.[16] The structure of the PR-S4 peptide complex (PDB 1 F7A) is shown on the right: PR, purple ribbon; the S4 peptide, grey sticks; and the raw MMM conformational ensemble for the nitroxide MTSL labels, blue sticks with the oxygen bearing the unpaired electron shown as red balls.

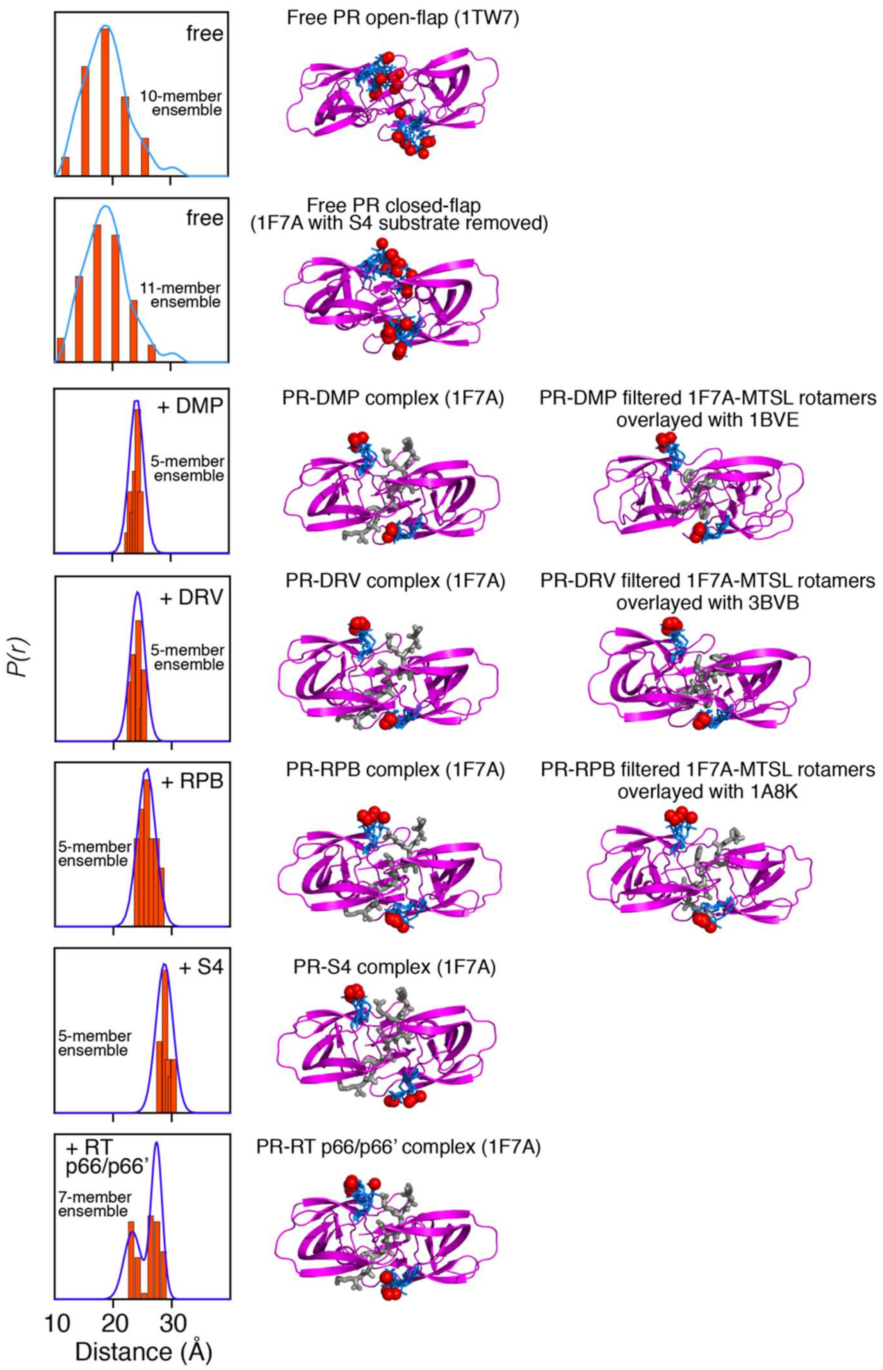

Figure S4. Comparison of experimental DEER-derived $P(r)$ distributions for nitroxide-labeled HIV-1 protease (V82C-R1) free and in the presence of drug inhibitors and peptides with the best-fit $P(r)$ distributions calculated from the atomic coordinates using the program RosettaEPR.[17] The color coding for the experimental DEER-derived $P(r)$ distributions is the same as that in Figs. S2 and S3: the experimental DEER-derived $P(r)$ distributions obtained by Tikhonov regularization using DeerAnalysis[12] are shown in light blue for free PR; for the complexes, a Gaussian(s) shown in dark blue was fitted to the peak(s) in the Tikhonov $P(r)$ distribution corresponding to the bound state. The RosettaEPR $P(r)$ distributions are displayed as histograms (bin width 1 Å, red), and the number of MTSL conformers (n) per subunit, giving rise to n^2 pairwise distances, in the final ensemble is indicated. As indicated in the SI methods section, all RosettaEPR calculations on the complexes were carried out using the coordinates of the PR-S4 complex (1F7A). The structures in the right-most column are the structures of the actual complexes with DMP, DRV and RPB onto which the MTSL rotamers calculated using the PR-S4 complex have been overlayed. Also note that for free PR (top two tows) the RosettaEPR $P(r)$ distributions obtained using the coordinates of both closed and open flap conformations of PR are almost identical.

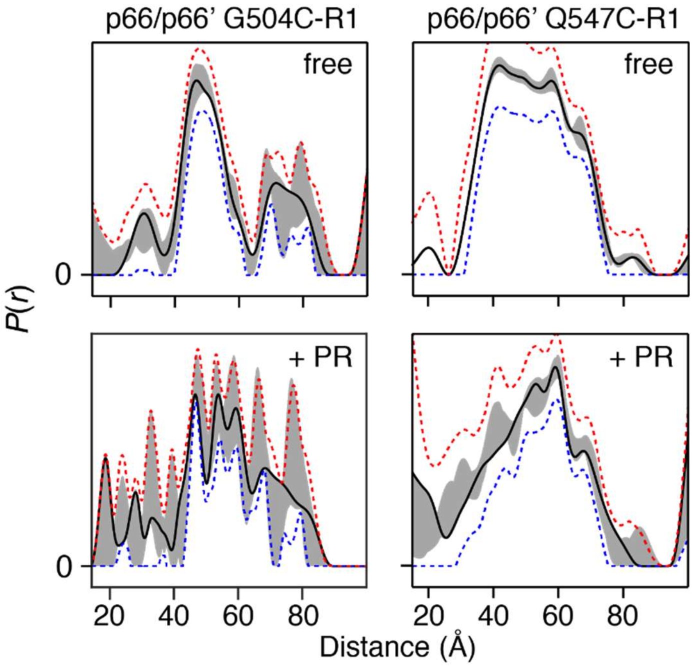

Figure S5. Results of DeerAnalysis[12] validation of *P*(*r*) distributions obtained using Tikhonov regularization for RT p66/p66' G504C-R1 and Q547-R1 free and in the presence of PR. The *P*(*r*) distribution with the best r.m.s.d. is displayed in black; the grey regions indicate the full *P*(*r*) variation over all trials; and the red and blue dotted lines are the upper (mean value probability plus two times its S.D.) and lower (mean value minus two times its S.D.) error estimates, respectively.

Supplementary references

[1] T. Schmidt, C. D. Schwieters, G. M. Clore, *Proc. Natl. Acad. Sci. U. S. A.* **2019**, *116*, 17809-17816.

[2] L. Deshmukh, J. M. Louis, R. Ghirlando, G. M. Clore, *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113*, 12456-12461.

[3] T. Schmidt, R. Ghirlando, J. Baber, G. M. Clore, *Chemphyschem* **2016**, *17*, 2987-2991.

[4] P. Y. Lam, P. K. Jadhav, C. J. Eyermann, C. N. Hodge, Y. Ru, L. T. Bacheler, J. L. Meek, M. J. Otto, M. M. Rayner, Y. N. Wong, *Science* **1994**, *263*, 380-384.

[5] J. M. Sayer, F. Liu, R. Ishima, I. T. Weber, J. M. Louis, *J. Biol. Chem.* **2008**, *283*, 13459-13470.

[6] A. D. Richards, L. H. Phylip, W. G. Farmerie, P. E. Scarborough, A. Alvarez, B. M. Dunn, P. H. Hirel, J. Konvalinka, P. Strop, L. Pavlickova, et al., *J. Biol. Chem.* **1990**, *265*, 7733-7736.

[7] R. Ishima, R. Ghirlando, J. Tozser, A. M. Gronenborn, D. A. Torchia, J. M. Louis, *J. Biol. Chem.* **2001**, *276*, 49110-49116.

[8] L. Polgar, Z. Szeltner, I. Boros, *Biochemistry* **1994**, *33*, 9351-9357.

[9] L. Deshmukh, V. Tugarinov, J. M. Louis, G. M. Clore, *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114*, E9855-E9862.

[10] M. Pannier, S. Veit, A. Godt, G. Jeschke, H. W. Spiess, *J. Magn. Reson.* **2011**, *213*, 316-325.

[11] S. Brandon, A. H. Beth, E. J. Hustedt, *J. Magn. Reson.* **2012**, *218*, 93-104.

[12] G. Jeschke, V. Chechik, P. Ionita, A. Godt, H. Zimmermann, J. Banham, C. R. Timmel, D. Hilger, H. Jung, *App. Magn. Reson.* **2006**, *30*, 473-498.

[13] M. Srivastava, E. R. Georgieva, J. H. Freed, *J Phys Chem A* **2017**, *121*, 2452-2465.

[14] M. Srivastava, J. H. Freed, *J. Phys. Chem. A* **2019**, *123*, 359-370.

[15] Y. Polyhach, E. Bordignon, G. Jeschke, *Phys. Chem. Chem. Phys.* **2011**, *13*, 2356-2366.

[16] G. Jeschke, *Protein Sci.* **2018**, *27*, 76-85.

[17] N. S. Alexander, R. A. Stein, H. A. Koteiche, K. W. Kaufmann, H. S. McHaourab, J. Meiler, *PLoS One* **2013**, *8*, e72851.

[18] K. M. Misura, D. Baker, *Proteins* **2005**, *59*, 15-29.

[19] P. Bradley, K. M. Misura, D. Baker, *Science* **2005**, *309*, 1868-1871.

[20] M. Prabu-Jeyabalan, E. Nalivaika, C. A. Schiffer, *Structure* **2002**, *10*, 369-381.

[21] J. Roche, J. M. Louis, A. Bax, *ChemBioChem* **2015**, *16*, 214-218.